

Weekly Report

1. 工作概述

本周我按照上周预定的计划开展工作

- (1) 我首先构建了一个训练数据, $D = [0,100]$ 它首先生成一个中心点 $X_c \in D^2$ 。然后随机生成一系列的点 $L = \{X_1, X_2, \dots, X_n\}$, 其中 $X_i \in D^4$, 和取值范围与 X_c 一致。我希望训练一个分类器使得 $(X_c, X_i, X_j) \rightarrow Y \in \{-1, 1\}$, 其中 $X_i, X_j \in L$, 当 $\|X_i - X_c\| < \|X_j - X_c\|$ (计算时只用到 X_i, X_j 的前两维) 时 $Y = 1$, 否则 $Y = -1$ 。
- (2) 在构建数据时我定义了一个偏序关系, 首先我根据 ListNet 的思路, 对一个列表 L , 通过数据学习 rank 规则, 输出一个 ranking, 取 TOP-1。我们的实验表明当 $n = 10$ 时, 可以达到 90% 的正确率; 当 $n = 20$ 时, 正确率就只有 55%; 当 $n = 40$ 时正确率仅有 20%
- (3) 由于 ListNet 当 n 急速增大时, 正确率显著下降, 为此我们考虑 Pairwise 的 Learning to rank 方法。我们首先选择的 Pairwise 的 learning to rank 方法名为 RankNet, 一种发表于 2005 年较为知名的 pair-wise learning to rank method。实验表明, 当 $n=50$ 时排序的正确率可达 90%。在该方法效果足够的情况下, 我们暂时不继续尝试其它的 learning to rank 方法

2. 下周计划

- (1) 把 RankNet 整合到 MapMatching 上, 用于提取局部特征
- (2) 完成论文初稿。可暂时先和局部匹配算法的结果做对比
- (3) 用 LSTM 或其它方法提取全局特征, 如果本周时间不够, 可以先不做

3. 论文阅读

《Learning to Rank Short Text Pairs with Convolutional Deep Neural Networks》

This paper uses CNN to feature engineering and NN to ranking.

《Adapting boosting for information retrieval measures》

The method proposed in this paper combines the strengths of boosted tree classification, and LambdaRank.

《Learning to Rank with Nonsmooth Cost Functions》

This method is also known as LambdaRank.

4. 时间安排

Date	Tasks	Duration
Mon.	Reading and Programming	9:00-11:30; 15:00-22:00
Tues. to Sun.	Reading and Programming	9:00-22:50

Work Time: above 50 hours